DESIGN AND ENGINEERING OF A MULTI-TARGET (MULTIPLEX) DNA SIMULANT TO EVALUATE NULCEIC ACID BASED ASSAYS FOR DETECTION OF BIOLOGICAL THREAT AGENTS

Mónica Carrera and José-Luis Sagripanti

Edgewood Chemical Biological Center (ECBC)
Research Development and Engineering Command,
Aberdeen Proving Ground, Maryland.

ABSTRACT

We designed and engineered a non-infectious Bio-threat simulant that included the nucleic acid signature of *Bacillus anthracis, Yersinia pestis, Coxiellla Burneti, Brucella sp., Francicella tularensis, Entherohemorragic E. coli, O157:H7, Burkholderia mallei, Burkholderia pseudomallei and Variola virus* (smallpox virus). A chimera of 2040 bp was engineered to produce PCR amplicons of different sizes in a single Multiplex reaction designed for the rapid identification of the threat agents selected above.

1. INTRODUCTION

1. 1. Significance and Impact of the Study

Nucleic acids-based technologies are a mainstay of DOD strategy to detect and identify biological threat agents. PCR amplification tests, in particular, have several advantages which include higher sensitivity and often lower cost than other approaches. However, most PCR methods target only one biological agent (amplifying only one primer pair at a time). Lack of standardized controls and protocols has contributed to the high rate of false positives and false alarms reported for PCR and other nucleic acid technologies. In addition, current biological simulants (*B. athrophaeus* [known before as B. *globigii*], *Erwinia herbicola* [renamed *Pantoameba agglomerans]*, and phage MS2) are particularly inadequate to evaluate specificity and sensitivity of nucleic acid-based tests, since the simulants do not share nucleic acid targets with any threat agent.

Using the actual bio-threat agents for testing is impractical since producing a number of different threat bacteria and viruses, isolating and characterizing them under adequate bio-containment, and preparing a representative control of each agent for test method evaluation represent nearly insurmountable logistic and economic difficulties. Therefore, our goal was to design and engineer a non-infectious simulant that included the nucleic acid signature of many bacterial and viral biological threat agents, within a single chimeric construct

1. 2. Background of the selective agents.

Bacillus anthracis is the etiological agent of anthrax and was the biological weapon used during the 2001 mail bioterrorist attacks. To date, several *B. anthracis* strains had been sequenced, but most are not available as full and annotated sequences. The only virulent strain of *B. anthracis* available in public databases is the "Ames ancestor" strain or A0581 strain. (Read et al, 2003)

Yersinia pestis, is the causative agent of the systemic invasive infectious disease classically referred to as "plague", and has been responsible for three devastating human pandemics separated by centuries. Due to the use by Japan during World War II and more recently to the identification of strains resistant to drugs (Galimand, M. et al, 1997), *Y. pestis* is an agent of biological warfare relevance.

Francisella tularensis is one of the most infectious pathogens known and is the etiological agent of tularemia, a disease of human and animals. Although this bacterium is nutritionally fastidious, it was developed as a weapon by Imperial Japan, the former Soviet Union, and the US. (Larsson, P. *et. al*, 2005). The sequenced strain corresponds to a fully virulent human isolate of *Francisella tularensis* subsp *tularensis* (strain SCHU S4, Larsson, P. el al, 2005)

Brucella species are etiological agents of brucellosis, a zoonotic disease endemic in many areas of the world, characterized by chronic infections in animals leading to abortion and infertility, and a systemic, febrile illness in humans. (Paulsen, I.T. et al 2002). *Brucella suis* was the first pathogenic organism weaponized by the US military during 1950s (Paulsen, I.T. et al, 2002). Since brucellosis threatens the food supply and causes undulant fever, a long, debilitating disease in humans, Brucella species are recognized as potential agricultural, civilian, and military bioterrorism agents.

Rickettsia are classified into two groups; the spotted fever group (SFG), which includes *R. conorii*, *R. sibirica*, and *R. rickettsii*, and the typhus group (TG), which includes *R. prowazekii* and *R. typhi*, according with the type of affection that they can cause. Both Japan, during World War II, and the former Soviet Union, during the

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 NOV 2006	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Design And Engineering Of A Multi-Target (Multiplex) Dna Simulant To Evaluate Nulceic Acid Based Assays For Detection Of Biological Threat Agents	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Edgewood Chemical Biological Center (ECBC) Research Development and Engineering Command, Aberdeen Proving Ground, Maryland.**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002075., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**8**	

Cold War, investigated the use of Rickettsiae as biological weapons. (McLead, M.P. et al 2004).

Two representatives of the *Burkholderia* genus with potential bio-warfare use have been completely sequenced, *B. mallei*, the etiologic agent of glanders, and *B. pseudomallei*, causative agent of melioidosis. A non-pathogenic specie, *B. thailandensis*, was also completely sequenced (Kim HS, *et al* 2005).

Coxiella burnetii, a highly virulent zoonotic pathogen and category B bioterrorism agent, was sequenced by the random shotgun method (Seshadri R. *et al* 2003)

Although the lifestyle and parasitic strategies of *C. burnetii* resemble that of Rickettsia*e* and Chlamydiae, their genome architectures differ considerably in terms of presence of mobile elements, extent of genome reduction, metabolic capabilities, and transporter profiles (Seshadri R. *et al* 2003)

Enterohemorrhagic Escherichia coli (EHEC) O157:H7 is a worldwide threat to public health and has been implicated in many outbreaks of hemorrhagic colitis, some of which included fatalities caused by hemolytic uremic syndrome (HUS). (Hayashi T. *et al*, 2001).

Variola virus, which causes smallpox, belongs to a genus of viruses known as Orthopoxvirus. Smallpox outbreaks involve either variola minor or the more deadly variola major.

2. MATERIALS AND METHODS

2 .1. Database and alignment of genomes

The genomes of many of the threat agents are public domain. All genomes used in this work were downloaded from NCBI (National Center for Biotechnology Information) (www.ncbi.nlm.nih.gov). The Basic Local Alignment Search Tool (BLAST, www.ncbi.nlm.nih.gov/BLAST) was used to find regions of local similarity between sequences. This program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST was used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

2. 2. Software and scripts.

The alignment of different strains were performed by ClustalX software (a windows interface to ClustalW multiple sequence alignment software) (Thompson, J.D et al 1997). All potential primers were generated by FastPCR, a program to design primers by Ruslan Kalendar (2006) "FastPCR, PCR primer design, DNA and protein tools, repeats and own database searches program" (www.biocenter.helsinki.fi/bi/programs/fastpcr.htm).

Several scripts were developed in Perl language to facilitate the analysis of the considerable amount of information that we generated during whole genome comparisons. Perl is a programming language that facilitates manipulation of strings (a set of consecutive characters) and has several modules specific for biological information handling (particularly BioPerl Project, www.bioperl.org).

3. RESULTS

3. 1. Search and download the available complete genome of each agent

For some of the agents, more than one complete genome is available. In those cases, all genomes were downloaded and used in some instance in this study.

Table 1. Complete bacterial genome sequences

Genome	Access numbers	Size (bp)
***B. anthracis* strain Ames**[1]	NC_003997	5,227,293
***B. anthracis* Ames "Ames ancestor"** [1]Read *et al* , 2003	NC_007322 pXO1 NC_007323 pXO2 NC_007530	181,677 94,830 5,227,419
***B. anthracis* strain Sterne** Okinaka *et al* , 1999	NC_005945 NC_001496 pXO1	5,228,663 181,654
***B. anthracis* strain Pasteur** Direct submission	NC_002146 pXO2	96,231
Brucella abortus strain 9-941 Halling *et al*, 2005	NC_006932 NC_006933	2,124,241 1,162,204
Brucella melitensis DelVecchio *et al* , 2002	NC_003317 NC_003318	2,117,144 1,177,787
Brucella abortus strain 2308 Chain *et al*, 2005	NC_007618 NC_007624	2,121,359 1,156,948
***Brucella suis* strain 1330** Paulsen *et al* , 2002	NC_004310 NC_004311	2,107,794 1,207,381
Francisella tularensis Larsson, P. *et al*, 2005	NC_006570	1,892,819
Rickettsia conorii Ogata *et al*, 2001	NC_003103	1,268,755
Rickettsia felis Ogata *et al*, 2005	NC_007109 NC_007110 NC_007111	1,485,148 62,829 39,263
Rickettsia prowazekii Andersson *et al*, 1998	NC_000963	1,111,523
Rickettsia typhi McLeod *et al*, 2004	NC_006142	1,111,496
***Yersinia pestis* CO92** Parkhill *et al*, 2001	NC_003131 NC_003132 NC_003134 NC_003143	70,305 9,612 96,210 4,653,728
***Yersinia pestis* KIM** Deng *et al* , 2002	NC_004088 NC_004838	4,600,755 100,990
***Yersinia pestis* 91001** Song *et al* , 2004	NC_005810 NC_005813 NC_005814 NC_005815 NC_005816	4,595,065 70,159 21,742 17,626 9,609
Yersinia pseudotuberculosis Chain *et al*, 2004	NC_006153 NC_006154 NC_006155	68,526 27,702 4,744,671
Burkholderia mallei Nierman *et al*, 2004	NC_006348 NC_006349	3,510,148 2,325,379
Burkholderia pseudomallei Holden *et al*, 2004	NC_006350 NC_006351	4,074,542 3,173,005
Burkholderia thailandensis Kim *et al*, 2005	NC_007650 NC_007651	2,914,771 3,809,201
***E. coli* O157 H7** Perna *et al*, 2001	NC_002127 NC_002128 NC_002695	3,306 92,721 5,498,450
***E. coli* O157 H7 EDL933** Makino *et al* , 1998 Hayashi *et al* , 2001	NC_002655	5,528,445
Coxiella burnetii Seshadri R. *et al* 2003	NC_002971	1,995,281

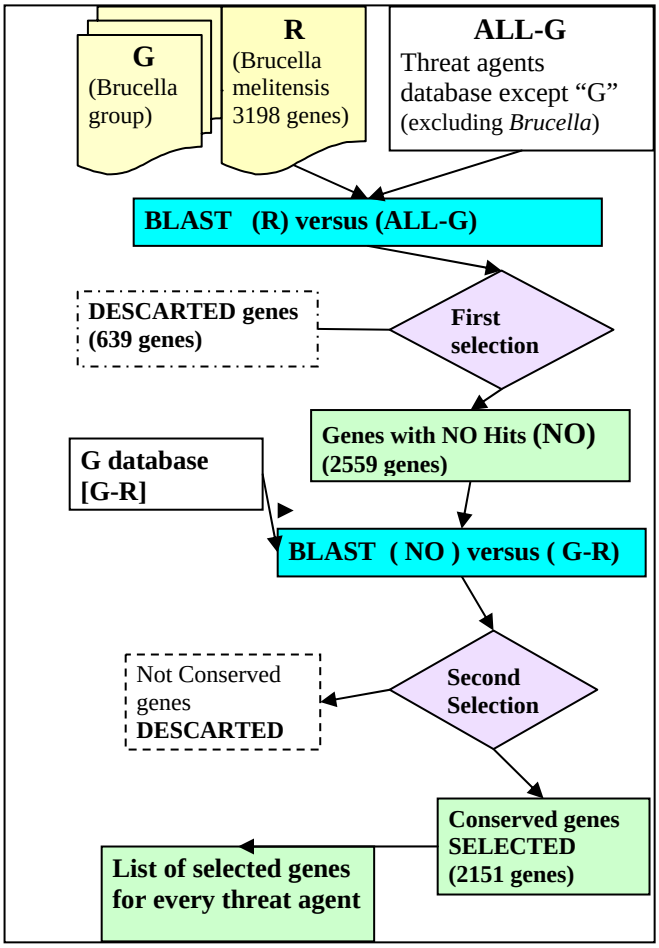
Because the availability of the complete genome for very closely related species, or even strains, comparisons within these groups of organisms were done separately, since levels of similarity are in a different order. Consequently, we selected one specie or strain as a "representative" of the group. The selection was made based on the importance of the threat to humans.

The complete list of genome sequences used is listed in **Table 1.** Additional genomes used for several comparisons were: *Bacillus cereus* ATCC 14579 (Ivanova N et al, 2003) and *Escherichia coli* K12 (Blattner FR et al 1997).

3. 1. 1. Database and genome comparison between microorganisms

For each particular threat agent, our goal was to identify specific gene sequences having two characteristic: a) to be absent in the other species listed in Table 1 and b) to be conserved within their own specie group.

Figure 1. Scheme of gene selection



G = group of a selected agent.
R = representative agent from a determined group
BLAST = comparison of sequences
In parenthesis we used *Brucella sp*. as an example.

Each threat genome was compared against all the other species genomes listed in Table 1 using BLAST as described in Materials and Methods. A systematic procedure for each individual gene was followed. Figure 1 shows a scheme representing all the steps that were performed using *Brucella sp.* as an example.

As shown on Figure 1, BLAST databases were created with all species genomes in Table 1 excluding the genomes of the specie group containing the agent in question. In the example shown in Figure 1, *Brucella melitensis*, is compared to all other species (*Bacillus anthracis, Yersinia, Coxiellla, Francicella, E. coli, Burkholderia and Variola virus*) listed in Table 1, but not to the other *Brucella* strains. This database was called ALL-G. The agent compared to all the rest of the species (*Brucella melitensis* in Figure 1) is called "representative agent"(R).

After the initial comparison with BLAST, (First Comparison in Figure 1) the resulting genes were grouped according to producing none, one, two, three, or more hits with the ALL-G database. A hit was considered a matching sequence between the "representative agent" with the genomes in the ALL-G database (with an error lower than 0.001). Alignment of at least 20-25 nucleotides were detected using these parameters. All genes that had some degree of similarity (more than one hit) were discarded and the genes with no hits were selected. These genes sequences specific for each threat organism were thus (negatively) selected for further analysis.

To select conserved genes within the same specie groups, a second comparison or BLAST was performed. This second alignment was done by creating an agent-specific database that included the complete genomes of all strains or specie within a group listed in Table 1 except the representative agent. Using Figure 1 example, *Brucella melitensis* (NO Hits) was compared against all strains in the *Brucella* group except *Brucella melitensis*. This new database was called G-R. Now the "representative agent" (R) was used as a query for a G-R database. The products of a positive selection in this comparison are the conserved genes within the different strains studied.

Our approach involving a two step analysis (consisting in a negative selection followed by positive selection) defined a set of genes conserved within closely related species or group (e.g. among all *B. anthracis* or among all *Brucella*) but with no sequence similarity with any of the others of the species groups listed in Table 1. Each group was analyzed separately taking into account the special characteristics that each of these different species have. Results from analysis of a few groups are described bellow as examples.

3. 1. .2. *Bacillus anthracis* group.

B. anthracis "Ames ancestor" was selected as the representative of this group, because it is fully virulent and the only strain of *B. anthracis* with both plasmids completely sequenced. The negative and positive selection analysis described above was then performed. *B. anthracis* "Ames ancestor" was used as the representative agent against the (ALL minus anthrax group, All-A) database. One-by-one all the genes in *B. anthracis* "Ames ancestor" were analyzed as described Figure 1. 208 genes that showed one or more hits with de complementary (All- A) database were discarded. A total of 5409 genes didn't show any hits, 204 corresponded to pXO1, 102 to pXO2, and 5103 to the chromosome. Interestingly, none of the genes of pXO1 and only 2 of pXO2 showed similarity with the other genomes studied. All the genes without hits were thus negatively selected for further analysis.

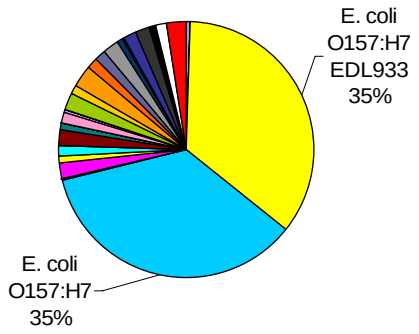
The negatively selected genes in the *B. anthracis* ancestor were analyzed and all genes that were not conserved among all other available *B. anthracis* (Ames, Sterna, and Pasteur) were discarded.

We found that all genes (204) were conserved from pXO1, as well as 102 genes from pXO2. In contrast, 342 genes were discarded from the chromosome because the gene sequences were not conserved among species. By negative and positive selection, a list of 4761 conserved genes conserved in the *Bacillus anthracis* group without any similarities with other threat organisms was obtained.

3. 1. 3. *Yersinia* group

In a similar approach to that described above, a list of genes conserved in the Yersinia group without any similarity with the other threat organisms was obtained. From a total of 4067 genes (including those in the chromosome and plasmids), 2262 genes did not show any hits with the ALL minus R database. We found that only 12 of the 170 total genes were conserved in the plasmids., We found that 1676 genes were conserved between species in the bacterial chromosome after discarding 416 genes The high degree of similarity founded could be caused by a shared common backbone between Yersinia and *E. coli*. Approximately 70% (3739 from a total of 5304 hits) corresponded to similarities with E. coli. (Figure2)

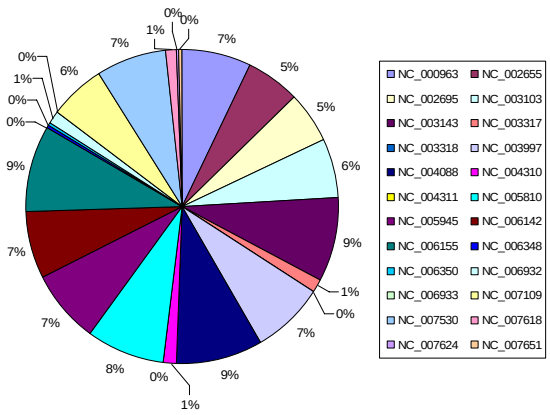
Figure 2. Distribution of hits in *Yersinia pestis* against other genomes



3. 1. .4. *Francisella tularensis.*

The complete genome of *Francisella tularensis* consists of a circular chromosome of 1,892,819 bp (NC_006570), with 1,603 predicted coding sequences (1,804 if pseudogenes are included). Following the same procedure used for *B. anthracis*, and represented in the scheme on Figure 1, "one by one" of each gene in the Francisella genome was compared against the "ALL minus Francicella" database. We found that there were no major similarities with any genome of other threat organisms but instead, the hits were distributed among several genomes in the database. (Figure 3). We found that 1420 out of 1603 total genes (88.6%) did not show any hits with the complementary database (All-Francicella) and only 183 genes were discarded based on similarities between *F. tularensis* and its complementary database. No further comparisons were done since there are not sequenced relatives of *Francisella tularensis* to search for group conserved sequences.

Figure 3. Hits distribution of *Francisella* genes against the complementary database

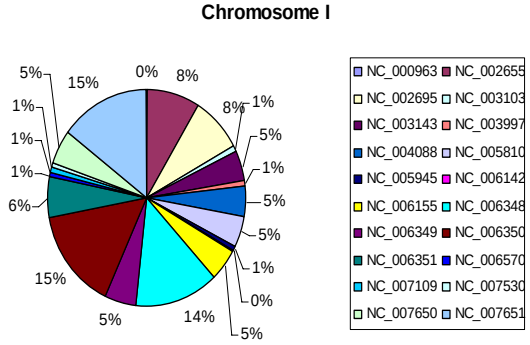


3. 1. 5. *Brucella* group

Based on the comparative genomics studies we chose *B. melitensis*, as a representative of the group since this organism shared a relatively larger number of genes with the other Brucella species. This allowed a better identification of common genes conserved among the Brucella group. A total of 639 genes were discarded (433 genes corresponded to chromosome I and 206 to chromosome II) after the First selection. A total of 2559 genes (1626 and 933 for Chromosome I and II respectively) did not show any similarities with the "ALL minus Brucella" database and therefore, were selected for further analysis.

Most of the hits in the genome of the Brucella group corresponded to genes in the *Burkholderia* genus (Figure 4). Sixty % and 65% of hits corresponding to chromosomes I and II of *Brucella*, respectively, were with genes belonging to the three *Burkholderia* genomes in the database *(B. mallei*, *B. pseudomallei* and *B. thailandensis*). This genetic similarity may be related to a common lifestyle shared between *Brucella* and *Burkholderia* (particularly *B. mallei*.) since organisms in both groups infect animals and are obligate parasites. Thus, these similarities could result from related genes associated with microbial survival. The similarities founded with *B. pseudomallei* could be related to the common backbone shared between Brucella and the Burkholderia genus, in spite of their differences in lifestyle, pathogenesis and genome content.

Figure 4. Hits distribution of Brucella genes against the complementary database.



Similar strategy as that described above was followed to analyze the Rickettsia group, Burkholderia genus, Escherichia group and *Coxiela burneti.*

Since the probability to find a specific DNA sequence absent in other organism is dramatically higher for bacterial genomes than for the smaller viral genomes, the analysis carried out with Variola virus (smallpox virus) genome differed from the approach indicated above. Conserved regions among all the 3 isolates of the Variola virus genome were selected by aligning the sequences using ClustalW (see Software and Scripts) algorithm for multiple sequence alignment.

3. 2. 1. Sizes selection

Once we had determined the specific target sequences in each selected microorganism, we established the size for each genome of the DNA fragment that would result by PCR amplification. An engineered chimera was designed to produce PCR amplicons of different sizes than the amplified fragments from the original pathogenic genome to identify false positives by knowing that simulant and pathogen should produce different size fragments.

Table 2 describes the sizes of the amplified products chosen for primer design. The indicated sizes were utilized as parameter for primer design using the FastPCR software. Two fragment sizes corresponding to each plasmid in *Bacillus anthracis* were selected because the absence of a plasmid in *B. anthracis* considerably reduces the pathogenicity. Thus, only strains or isolates carrying both plasmids are fully virulent. Therefore, the identification of virulent isolates of *B. anthracis* must be done by detecting both plasmids.

Table 2. Selected sizes for pathogenic microorganism and simulant amplified fragments

Organism or group	Preferred size in pathogen	Size in simulant
Bacillus anthracis pXO1	150	205
Bacillus anthracis pXO2	169	220
Yersinia group	200	235
Francisella tularensis	230	100
Burkholderia group	260	115
Rickettsia group	290	130
Coxiella burnetti	310	145
Brucella group	330	160
Escherichia coli O157:H7 group	350	175
Variola virus	380	190

3. 2. .2. Primer design

Primers 22-26 nucleotides long were designed with an annealing temperature above 55ºC and a PCR product with the desired length indicated in Table 2 by using the FastPCR software as indicated in Materials and Methods. To generate a more extensive potential primer pair list, the amplified size parameter used was within a range of ±20 nucleotides of the selected sequence. All the remaining parameter settings were the default of the software. The whole gene sequences of the selected bacteria genes were used for primer design. All the

possible primers were predicted for each DNA sequence selected. Then, a list of all the possible "primer pairs" able to generate an amplified DNA fragment of the expected length was generated. A Microsoft Excel file containing all the primers and primer pairs was generated for each selected gene as output from FastPCR. (Data not shown).

3. 2. 3. Further selection of primer pairs.

Possible yet unspecific primers (able to bind to non-related genomes in Table 1) were discarded by a preliminary selection step. All primers were subjected to an "in silico" PCR prediction using FastPCR. Those primers that showed more than 80% similarity and 5 matches in the 3´end of the last 7 bases generating an amplified fragment in any genome were discarded. Using a Perl script specifically designed for this purpose, we made a list of primers for the selected genes that showed 100% similarities with the target genome and a similarity lower than 80% with any other genome in this study.

3. 2. 4. Multiplex design

After identifying a considerable number of potential primers pairs, we focused on the generation of primer groups to build the chimeric positive control and test all threat organisms in an in silico multiplex reaction. A primer pair for each genome fragment was selected from the primer pair list constructed with the Pearl script indicated above based on these following criteria

1) Preferably primers length of 26 bp
2) Quality value of the primers (high)
3) Similar annealing temperature among the group of primers
4) Theoretical amplified fragment size closest to that indicated in Table 2.

This criterion allowed creating several primer groups. The groups were tested in two different ways for their use in a multiplex reaction. First, we did the FastPCR function "List of primers to test" that check for dimer formation among the group and second, we did an in silico PCR against each genome.

3. 3. Simulant assembly

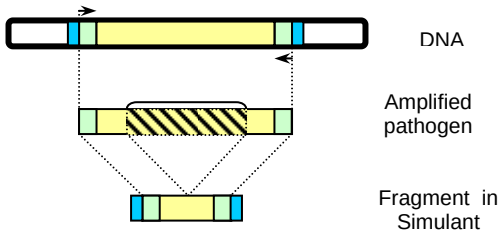
3. 3. 1. In silico test for multiplex group

The final test was to perform in silico PCR against each genome assuring that only the desired fragment was present in the corresponding genome and none (or unlikely) unspecific fragments appeared. To this purpose, fragments of several kb in length with primer similarity to other genomes below 80% were considered acceptable. The best choice of primers for multiplex PCR was finally selected after repeated analysis of several groups of primers, manual inspection of the output, and replacement of those primers that performed poorly.

3. 3. .2. Design of fragment for each genome

After obtaining the primers and amplified fragments for each genome, the chimerical molecule to be used as simulant in PCR reactions was designed. This molecule is being synthesized. The length of simulant amplified fragments differed from those in actual genomes, as detailed in Table 2. The fragments of the sizes indicated in Table 2 were obtained by deleting bases in the middle of the amplified sequences. At each side of the selected primers were added the 10 base-long flanking sequences present in the original genome. In this way, primers designed over approximately 40 bp around the primer selected could used in case of experimental need (Figure 5).

Figure 5. Scheme showing the design used for each fragment.



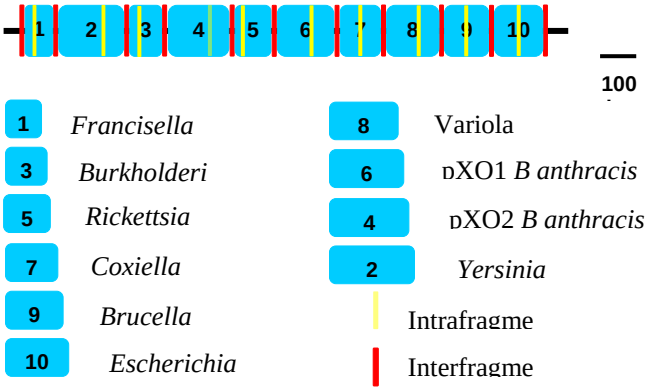
To each fragment we added two restriction sites in the middle of the sequence (EcoRI –GAATTC– and SmaI –CCCGGG–). These enzymes do not cut any of the amplified fragments from any of the genomes of interest. Therefore these two enzymes could be used to digest these fragments in case of false positive results or suspected contamination.

3. 3. .3. Chimera design and assembly

After design of all the fragments for each genome, the selected fragments in a chimerical molecule were joined. Between each of the fragments in the chimera, two additional restriction sites were added to perform a digestion step before the amplification process. This step ensures that no fragments longer than expected would be produced. This digestion was necessary, since the amplification of two consecutives fragments by primers between his extremes could possibly confound results. Thus, the specific sites for the enzymes BamHI (–GGATCC–) and HindIII (–AAGCTT–) were introduced between each fragment and also at beginning and end of the chimerical molecule.

A scheme showing this organization and the resulting chimera is showed in Figure 6.

Figure 6 A scheme showing the organization and the resulting chimera



4. CONCLUSIONS

The multiplex simulant molecule engineered here could be used to spike samples and afterward evaluate the performance of nucleic acid-based bio-detectors and diagnostic products of interest in biodefense. The proposed multiplex simulant would reduce the need of using individual bio-threat agents or their DNA as positive controls. Thus, the multiplex simulant could be used to test military detectors without exposing testers or trainees to pathogenic biological agents. In addition, a single standard multiplex simulant could be issued as positive control to evaluate and monitor nucleic acid-based biological testing platforms, including novel sensors and detectors. This multiplex simulant could be used to compare the performance of a variety of technologies used or envisioned in Biodefense. Easier, cheaper, and improved evaluation of technologies should assure continued reliability of biological detectors and reduced false alarms which degrade operational capabilities by unnecessary masking and gowning.

5. REFERENCES

Andersson SG, *et al*. (1998). The genome sequence of Rickettsia prowazekii and the origin of mitochondria. Nature. 1998 Nov 12; 396(6707):133-40.

Blattner FR, *et al*. (1997). The complete genome sequence of Escherichia coli K-12. Science. 1997 Sep 5;277(5331):1453-74.

Chain PS, *et al*. (2004) Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis. Proc Natl Acad Sci U S A. 2004 Sep 21; 101(38):13826-31

Chain PS, *et al*. (2005). Whole-genome analyses of speciation events in pathogenic Brucellae. Infect Immun. 2005 Dec; 73(12):8353-61

DelVecchio VG, *et al*. (2002). The genome sequence of the facultative intracellular pathogen Brucella melitensis. Proc Natl Acad Sci U S A. 2002 Jan 8; 99(1):443-8.

Deng W, *et al*. (2002). Genome sequence of Yersinia pestis KIM. J Bacteriol. 2002 Aug; 184(16):4601-11

Galimand M, *et al*. (1997). Multi drug resistance in Yersinia pestis mediated by a transferable plasmid.(1997) N Engl J Med. Sep 4;337(10):677-80.

Halling SM, *et al*. (2005).Completion of the genome sequence of Brucella abortus and comparison to the highly similar genomes of Brucella melitensis and Brucella suis. J Bacteriol. 2005 Apr;187(8):2715-26

Hayashi T, *et al*. (2001). Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. 2001 Feb 28;8(1):11-22.

Holden MT, *et al*. (2004) Genomic plasticity of the causative agent of melioidosis, Burkholderia pseudomallei. Proc Natl Acad Sci U S A. 2004 Sep 28;101(39):14240-5.

Ivanova N, *et al*.(2003). Genome sequence of Bacillus cereus and comparative analysis with Bacillus anthracis. Nature. 2003 May 1; 423(6935):87-91

Kim HS, *et al*. (2005) Bacterial genome adaptation to niches: divergence of the potential virulence genes in three Burkholderia species of different survival strategies. BMC Genomics. 2005 Dec 7;6:174.

Larsson P, *et al*. (2005). The complete genome sequence of Francisella tularensis, the causative agent of tularemia. Nat Genet. 2005 Feb; 37(2):153-9.

Makino K, *et al*. (1998). Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic Escherichia coli O157:H7 derived from Sakai outbreak. DNA Res. 1998 Feb 28;5(1):1-9

McLeod MP, *et al*.(2004) Complete genome sequence of Rickettsia typhi and comparison with sequences of other rickettsiae. J Bacteriol. 2004 Sep; 186(17):5842-55

Nierman WC, *et al*. (2004) Structural flexibility in the Burkholderia mallei genome. Proc Natl Acad Sci U S A. 2004 Sep 28; 101(39):14246-51

Ogata H, *et al*. (2001). Mechanisms of evolution in Rickettsia conorii and R. prowazekii. Science. 2001 Sep 14; 293(5537):2093-8.

Ogata H, *et al*. (2005). The genome sequence of Rickettsia felis identifies the first putative conjugative plasmid in an obligate intracellular parasite. PLoS Biol. 2005 Aug;3(8):e248.

Okinaka R, *et al*. (1999) Sequence, assembly and analysis of pX01 and pX02. J Appl Microbiol. Aug;87(2):261-2.

Parkhill J, *et al*. (2001) Genome sequence of Yersinia pestis, the causative agent of plague. Nature. 2001 Oct 4; 413(6855):523-7

Paulsen IT, *et al*.(2003) Complete genome sequence of the Q-fever pathogen Coxiella burnetii. Proc Natl

Acad Sci U S A. 2003 Apr 29; 100(9):5455-60. Epub 2003 Apr 18.
Perna NT, *et al* (2001) Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. Nature Jan 25; 409(6819):529-33.
Read TD, *et al*. (2003). The genome sequence of Bacillus anthracis Ames and comparison to closely related bacteria. Nature. 2003 May 1; 423(6935):81-6.
Shchelkunov SN, *et al*. (1994) Analysis of the nucleotide sequence of 53 kbp from the right terminus of the genome of variola major virus strain India-1967. Virus Res. 1994 Dec;34(3):207-36.
Song Y, *et al*.(2004) Complete genome sequence of Yersinia pestis strain 91001, an isolate avirulent to humans DNA Res. Jun 30;11(3):179-97.